

# Italian vs. Mexican Food

The below script provides an analytic approach for assessing American preferences of Italian vs. Mexican food. Using data from the US Census and the Yelp API, the script randomly selects over 500 zip codes and aggregates Yelp review s from the 20 most popular Italian and Mexican restaurants in each area. The data is then parsed and analyzed using Python Pandas and Matplotlib.

```
# Dependencies
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import requests
import time
import json
import seaborn
from scipy.stats import ttest_ind

# Yelp API Key (Keys Hidden)
ykey_id = "XXX"
ykey_secret = "XXX"
ykey_access_token = "g16k6JmewUhzjMVBv0I2x4Bz_NRiEggSqj1GbTaejmbzvBJXgI36FPgWq8nEL9QQ6wU5H4h41dxPkxVjHf1awtH69m1kcXQuHev5PuWBTcdBEAbdJR0HN13d4tpwXYx"
```

## Zip Code Sampling

```
# Import the census data into a Pandas DataFrame
census_pd = pd.read_csv("Census_Data.csv")

# Preview the data
census_pd.head()
```

	Zipcode	Address	Population	Median Age	Household Income	Per Capita Income	
0	15081	South Heights, PA 15081, USA	342	50.2	31500.0	22177	
1	20615	Broomes Island, MD 20615, USA	424	43.4	114375.0	43920	
2	50201	Nevada, IA 50201, USA	8139	40.4	56619.0	28908	
3	84020	Draper, UT 84020, USA	42751	30.4	89922.0	33164	
4	39097	Louise, MS 39097, USA	495	58.0	26838.0	17399	

```
# Select all zip codes with populations over 1000 from a pre-set list of 700 randomly selected zip code locations
selected_zips = census_pd.sample(n=700)
selected_zips = selected_zips[selected_zips["Population"].astype(int) > 1000]

# Visualize
selected_zips.head()
```

	Zipcode	Address	Population	Median Age	Household Income	Per Capita Income	
117	76556	Milano, TX 76556, USA	1392	50.7	39375.0	21201	
135	72039	Damascus, AR 72039, USA	2402	40.8	33857.0	19838	
389	61606	Peoria, IL 61606, USA	7989	21.9	35904.0	17917	
270	47232	Elizabethtown, IN 47232, USA	3280	37.4	60128.0	21838	
67	60565	Naperville, IL 60565, USA	40864	40.8	113581.0	45408	

```
# Show the total number of zip codes that met our population cut-off
selected_zips.count()
```

Zipcode	524
Address	524
Population	524
Median Age	524

```
Household Income    523
Per Capita Income    524
dtype: int64
```

```
# Show the average population of our representative sample set
selected_zips["Population"].mean()
```

```
13949.314885496184
```

```
# Show the average population of our representative sample set
selected_zips["Household Income"].mean()
```

```
56315.009560229446
```

```
# Show the average population of our representative sample set
selected_zips["Median Age"].mean()
```

```
39.937977099236626
```

## Yelp Data Retrieval

```
# Create Two DataFrames to store the Italian and Mexican Data
italian_data = pd.DataFrame();
mexican_data = pd.DataFrame();

# Setup the DataFrames to have appropriate columns
italian_data["Zip Code"] = ""
italian_data["Italian Review Count"] = ""
italian_data["Italian Average Rating"] = ""
italian_data["Italian Weighted Rating"] = ""

mexican_data["Zip Code"] = ""
mexican_data["Mexican Review Count"] = ""
mexican_data["Mexican Average Rating"] = ""
mexican_data["Mexican Weighted Rating"] = ""

# Include Yelp Token
headers = {"Authorization": "Bearer gl6k6JmewUhzyMVBv0I2x4Bz_NRIEggSqj1GbTaejmbzvBJXgI36FPgWoq8nEL9QQ6wU5H4h41dxPkxVjHf1awtH69m1kcXQuHev5PuWbtcdBEAbdJR0HN"}
counter = 0

# Loop through every zip code
for index, row in selected_zips.iterrows():

    # Add to counter
    counter = counter + 1

    # Create two endpoint URLs:
    target_url_italian = "https://api.yelp.com/v3/businesses/search?term=Italian&location=%s" % (row["Zipcode"])
    target_url_mexican = "https://api.yelp.com/v3/businesses/search?term=Mexican&location=%s" % (row["Zipcode"])

    # Print the URLs to ensure logging
    print(counter)
    print(target_url_italian)
    print(target_url_mexican)

    # Get the Yelp Reviews
    yelp_reviews_italian = requests.get(target_url_italian, headers=headers).json()
    yelp_reviews_mexican = requests.get(target_url_mexican, headers=headers).json()

    # Calculate the total reviews and weighted rankings
    italian_review_count = 0
    italian_weighted_review = 0
    mexican_review_count = 0
    mexican_weighted_review = 0

    # Use Try-Except to handle errors
    try:

        # Loop through all records to calculate the review count and weighted review value
        for business in yelp_reviews_italian["businesses"]:

            italian_review_count = italian_review_count + business["review_count"]
```

```

        italian_weighted_review = italian_weighted_review + business["review_count"] * business["rating"]

for business in yelp_reviews_mexican["businesses"]:
    mexican_review_count = mexican_review_count + business["review_count"]
    mexican_weighted_review = mexican_weighted_review + business["review_count"] * business["rating"]

# Append the data to the appropriate column of the data frames
italian_data.set_value(index, "Zip Code", row["Zipcode"])
italian_data.set_value(index, "Italian Review Count", italian_review_count)
italian_data.set_value(index, "Italian Average Rating", italian_weighted_review / italian_review_count)
italian_data.set_value(index, "Italian Weighted Rating", italian_weighted_review)

mexican_data.set_value(index, "Zip Code", row["Zipcode"])
mexican_data.set_value(index, "Mexican Review Count", mexican_review_count)
mexican_data.set_value(index, "Mexican Average Rating", mexican_weighted_review / mexican_review_count)
mexican_data.set_value(index, "Mexican Weighted Rating", mexican_weighted_review)

except:
    print("Uh oh")

```

```

1
https://api.yelp.com/v3/businesses/search?term=Italian&location=76556
https://api.yelp.com/v3/businesses/search?term=Mexican&location=76556
2
https://api.yelp.com/v3/businesses/search?term=Italian&location=72039
https://api.yelp.com/v3/businesses/search?term=Mexican&location=72039
3
https://api.yelp.com/v3/businesses/search?term=Italian&location=61606
https://api.yelp.com/v3/businesses/search?term=Mexican&location=61606
4
https://api.yelp.com/v3/businesses/search?term=Italian&location=47232
https://api.yelp.com/v3/businesses/search?term=Mexican&location=47232
5
https://api.yelp.com/v3/businesses/search?term=Italian&location=60565
https://api.yelp.com/v3/businesses/search?term=Mexican&location=60565
6
https://api.yelp.com/v3/businesses/search?term=Italian&location=20634
https://api.yelp.com/v3/businesses/search?term=Mexican&location=20634
7
https://api.yelp.com/v3/businesses/search?term=Italian&location=71046
https://api.yelp.com/v3/businesses/search?term=Mexican&location=71046
8
https://api.yelp.com/v3/businesses/search?term=Italian&location=76950
https://api.yelp.com/v3/businesses/search?term=Mexican&location=76950
9
https://api.yelp.com/v3/businesses/search?term=Italian&location=66507
https://api.yelp.com/v3/businesses/search?term=Mexican&location=66507
10
https://api.yelp.com/v3/businesses/search?term=Italian&location=37923
https://api.yelp.com/v3/businesses/search?term=Mexican&location=37923
11
https://api.yelp.com/v3/businesses/search?term=Italian&location=30268
https://api.yelp.com/v3/businesses/search?term=Mexican&location=30268
12
https://api.yelp.com/v3/businesses/search?term=Italian&location=64081
https://api.yelp.com/v3/businesses/search?term=Mexican&location=64081
13
https://api.yelp.com/v3/businesses/search?term=Italian&location=35117
https://api.yelp.com/v3/businesses/search?term=Mexican&location=35117
14
https://api.yelp.com/v3/businesses/search?term=Italian&location=16930
https://api.yelp.com/v3/businesses/search?term=Mexican&location=16930
15
https://api.yelp.com/v3/businesses/search?term=Italian&location=55735
https://api.yelp.com/v3/businesses/search?term=Mexican&location=55735
16
https://api.yelp.com/v3/businesses/search?term=Italian&location=50315
https://api.yelp.com/v3/businesses/search?term=Mexican&location=50315
17
https://api.yelp.com/v3/businesses/search?term=Italian&location=62330
https://api.yelp.com/v3/businesses/search?term=Mexican&location=62330
18
https://api.yelp.com/v3/businesses/search?term=Italian&location=4357
https://api.yelp.com/v3/businesses/search?term=Mexican&location=4357
19
https://api.yelp.com/v3/businesses/search?term=Italian&location=35801
https://api.yelp.com/v3/businesses/search?term=Mexican&location=35801
20
https://api.yelp.com/v3/businesses/search?term=Italian&location=63039
https://api.yelp.com/v3/businesses/search?term=Mexican&location=63039
21
https://api.yelp.com/v3/businesses/search?term=Italian&location=71353
https://api.yelp.com/v3/businesses/search?term=Mexican&location=71353
22
https://api.yelp.com/v3/businesses/search?term=Italian&location=7724
https://api.yelp.com/v3/businesses/search?term=Mexican&location=7724
23
https://api.yelp.com/v3/businesses/search?term=Italian&location=27283
https://api.yelp.com/v3/businesses/search?term=Mexican&location=27283

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

```
518
https://api.yelp.com/v3/businesses/search?term=Italian&location=95051
https://api.yelp.com/v3/businesses/search?term=Mexican&location=95051
519
https://api.yelp.com/v3/businesses/search?term=Italian&location=97321
https://api.yelp.com/v3/businesses/search?term=Mexican&location=97321
520
https://api.yelp.com/v3/businesses/search?term=Italian&location=5483
https://api.yelp.com/v3/businesses/search?term=Mexican&location=5483
521
https://api.yelp.com/v3/businesses/search?term=Italian&location=60468
https://api.yelp.com/v3/businesses/search?term=Mexican&location=60468
522
https://api.yelp.com/v3/businesses/search?term=Italian&location=50201
https://api.yelp.com/v3/businesses/search?term=Mexican&location=50201
523
https://api.yelp.com/v3/businesses/search?term=Italian&location=73951
https://api.yelp.com/v3/businesses/search?term=Mexican&location=73951
524
https://api.yelp.com/v3/businesses/search?term=Italian&location=42031
https://api.yelp.com/v3/businesses/search?term=Mexican&location=42031
```

```
# Preview Italian Data
italian_data.head()
```

	Zip Code	Italian Review Count	Italian Average Rating	Italian Weighted Rating	
117	76556	63	3.78571	238.5	
135	72039	266	3.81955	1016	
389	61606	66	3.2197	212.5	
270	47232	420	3.77857	1587	
67	60565	2829	3.92824	11113	

```
# Preview Mexican Data
mexican_data.head()
```

	Zip Code	Mexican Review Count	Mexican Average Rating	Mexican Weighted Rating	
117	76556	97	4.1134	399	
135	72039	256	4.11133	1052.5	
389	61606	378	3.64286	1377	
270	47232	222	4.16892	925.5	
67	60565	2842	3.94053	11199	

## Calculate Summaries

```
# Total Mexican Reviews
mexican_data["Mexican Review Count"].sum()
```

```
476889
```

```
# Total Italian Reviews
italian_data["Italian Review Count"].sum()
```

```
573733
```

```
# Average Mexican Rating
mexican_data["Mexican Weighted Rating"].sum() / mexican_data["Mexican Review Count"].sum()
```

```
3.909732663156416
```

```
# Average Italian Rating
italian_data["Italian Weighted Rating"].sum() / italian_data["Italian Review Count"].sum()
```

```
3.9446641556263975
```

```
# Combine DataFrames into a single DataFrame
combined_data = pd.merge(mexican_data, italian_data, on="Zip Code")
combined_data.head()
```

	Zip Code	Mexican Review Count	Mexican Average Rating	Mexican Weighted Rating	Italian Review Count	Italian Average Rating	Italian Weighted Rating
0	76556	97	4.1134	399	63	3.78571	238.5
1	72039	256	4.11133	1052.5	266	3.81955	1016
2	61606	378	3.64286	1377	66	3.2197	212.5
3	47232	222	4.16892	925.5	420	3.77857	1587
4	60565	2842	3.94053	11199	2829	3.92824	11113

```
# Determine Total Review Count and Rating "Wins" by City (Winner Take All)
combined_data["Rating Wins"] = np.where(combined_data["Mexican Average Rating"] > combined_data["Italian Average Rating"], "Mexican", "Italian")
combined_data["Review Count Wins"] = np.where(combined_data["Mexican Review Count"] > combined_data["Italian Review Count"], "Mexican", "Italian")
```

```
# View Combined Data
combined_data.head()
```

	Zip Code	Mexican Review Count	Mexican Average Rating	Mexican Weighted Rating	Italian Review Count	Italian Average Rating	Italian Weighted Rating	Rating Wins	Review Count Wins
0	76556	97	4.1134	399	63	3.78571	238.5	Mexican	Mexican
1	72039	256	4.11133	1052.5	266	3.81955	1016	Mexican	Italian
2	61606	378	3.64286	1377	66	3.2197	212.5	Mexican	Mexican
3	47232	222	4.16892	925.5	420	3.77857	1587	Mexican	Italian
4	60565	2842	3.94053	11199	2829	3.92824	11113	Mexican	Mexican

```
# Tally number of cities where one food variety wins on ratings over the other
combined_data["Rating Wins"].value_counts()
```

```
Mexican    273
Italian    245
Name: Rating Wins, dtype: int64
```

```
# Tally number of cities where one food variety wins on review counts over the other
combined_data["Review Count Wins"].value_counts()
```

```
Italian    298
Mexican    220
```

Name: Review Count Wins, dtype: int64

## Display Summary of Results

```
# Model 1: Head-to-Head Review Counts
italian_summary = pd.DataFrame({"Review Counts": italian_data["Italian Review Count"].sum(),
                                "Rating Average": italian_data["Italian Average Rating"].mean(),
                                "Review Count Wins": combined_data["Review Count Wins"].value_counts()["Italian"],
                                "Rating Wins": combined_data["Rating Wins"].value_counts()["Italian"]}, index=["Italian"])

mexican_summary = pd.DataFrame({"Review Counts": mexican_data["Mexican Review Count"].sum(),
                                "Rating Average": mexican_data["Mexican Average Rating"].mean(),
                                "Review Count Wins": combined_data["Review Count Wins"].value_counts()["Mexican"],
                                "Rating Wins": combined_data["Rating Wins"].value_counts()["Mexican"]}, index=["Mexican"])

final_summary = pd.concat([mexican_summary, italian_summary])
final_summary
```

	Rating Average	Rating Wins	Review Count Wins	Review Counts
Mexican	3.826588	273	220	476889
Italian	3.806869	245	298	573733

```
# Plot Rating Average
plt.clf()
final_summary["Rating Average"].plot.bar()
plt.title("Yelp Ratings by Food Variety")
plt.ylabel("Average Rating")
plt.xlabel("Food Variety")
plt.xticks(rotation="horizontal")
plt.grid(True)
plt.show()
```

png

```
# Plot Rating Wins
plt.clf()
final_summary["Rating Wins"].plot.bar()
plt.title("# of Zip Codes with Preference by Food Variety According to Rating")
plt.ylabel("Number of Zip Codes")
plt.xlabel("Food Variety")
plt.xticks(rotation="horizontal")
plt.grid(True)
plt.show()
```

png

```
# Plot Review Count
plt.clf()
final_summary["Review Counts"].plot.bar()
plt.title("Yelp Review Counts by Food Variety")
plt.ylabel("Review Counts")
plt.xlabel("Food Variety")
plt.xticks(rotation="horizontal")
plt.grid(True)
plt.show()
```

png

```
# Plot Review Count
plt.clf()
final_summary["Review Count Wins"].plot.bar()
plt.title("# of Zip Codes with Preference by Food Variety According to Review Counts")
plt.ylabel("Number of Zip Codes")
plt.xlabel("Food Variety")
plt.xticks(rotation="horizontal")
plt.grid(True)
plt.show()
```

png

```
# Histogram Italian Food (Ratings)
plt.figure()

# Subplot 1 (Italian)
plt.subplot(121)
combined_data["Italian Average Rating"].plot.hist(bins=[0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5.0], color="blue", alpha=0.6)
plt.xlabel("Italian Restaurant Ratings")
plt.xlim([1, 5.0])
plt.ylim([0, 400])

# Subplot 2 (Mexican)
plt.subplot(122)
combined_data["Mexican Average Rating"].plot.hist(bins=[0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5.0], color="red", alpha=0.6)
plt.xlabel("Mexican Restaurant Ratings")
plt.xlim([1, 5.0])
plt.ylim([0, 400])

# Show Plot
plt.show()
```

png

## Statistical Analysis

```
# Run a t-test on average rating and number of reviewers
mexican_ratings = combined_data["Mexican Average Rating"]
italian_ratings = combined_data["Italian Average Rating"]

mexican_review_counts = combined_data["Mexican Review Count"]
italian_review_counts = combined_data["Italian Review Count"]
```

```
# Run T-Test on Ratings
ttest_ind(mexican_ratings.values, italian_ratings.values, equal_var=False)
```

```
Ttest_indResult(statistic=1.0702777109866186, pvalue=0.28474610492878616)
```

```
# Run T-Test on Review Counts
ttest_ind(mexican_review_counts.values, italian_review_counts.values, equal_var=False)
```

```
Ttest_indResult(statistic=-1.9031271166792225, pvalue=0.057326353522100949)
```

## Conclusions

Based on our analysis, it is clear that the American preference for Italian and Mexican food are similar in nature. As a whole, Americans rate Mexican and Italian restaurants at statistically similar scores (Avg score: 3.8, p-value: 0.285). However, there exists statistically significant evidence that Americans write more reviews of Italian restaurants than Mexican. This may indicate that there is an increased interest in visiting Italian restaurants at an experiential level. However, it may also merely suggest that Yelp users enjoy writing reviews on Italian restaurants more than Mexican restaurants.